

修 士 論 文 の 和 文 要 旨

研究科・専攻	大学院 情報システム 学研究科 社会知能情報学 専攻 博士前期課程		
氏 名	奥地 大容	学籍番号	0751004
論 文 題 目	講義音声から抽出した重要語を用いたアノテーション付与に関する研究		
<p>要 旨</p> <p>講義ビデオにアノテーションを付与する上で、問題点が大きく分けて2つある。1つは、量の問題で、学習者にとって適切な量のアノテーションを推定することは困難であること。もう1つは作業量が膨大で人的コストがかかる点である。また、講義動画は動画部分では大まかな流れを示し、具体的な内容は音声で伝えるものが一般的であるが音声部分に情報が偏り過ぎる傾向がある。特に近年ではスライド教材を用いた講義が増えているため、全体としての流れは把握できても具体的な内容を理解するのは難しくなっている。</p> <p>本研究では講義ビデオの各シーンに教授者の重要発話をアノテーションとして半自動的に付与するシステムを構築し問題の解決をはかり、実験と評価を行った。本システムでは、①スライドから主題となる単語の抽出、②重要発話の抽出を行う。</p> <p>まず①では、文書から重要語を抽出する方法として tf-idf 法などが存在するがそれらの方法はある程度の大きさをもった文書が解析の対象とされている。しかし、本研究で解析の対象としている発表スライドは一般的に含まれる単語数が少ないため、他の重要語の抽出が必要となる。そこで本研究ではスライドからの重要語抽出のために、語の接続に着目した。</p> <p>②では、アノテーションを行う際にスライドのような箇条書きでは効果が低いため、本研究では教授者から文章として記述できる重要な発話を抽出する。①から取得した情報との共起頻度を計算することにより、スライドにない情報でも講義の主題とリンクした重要発話の抽出を可能とした。また、講義ビデオでは流れを重視した発話がなされることが多いため、本研究では接続表現にも注目し、特定の機能を持つ接続表現が含まれている発話に重みづけもおこなっている。</p> <p>実験の結果、教授者がスライドに記述されていないが重要性のある情報を多く発話している場面や、図や表が中心で文字情報の少ないスライドの場面では特に高い評価を得ることができ、本システムの有効性を確認できた。</p>			